StreamFinancial

28th February 2014

Financial Stability Board
CH-4002
Basel
Switzerland

Dear Sir/Madam:

Stream Financial welcomes the opportunity to share with the Financial Stability Board (FSB) our views on the consultative document on: Feasibility study on approaches to aggregate OTC derivatives data. We hope that you will find our contribution helpful and welcome any questions you may have.

In our opinion, given the current data and technology options and early stage development of the industry standards and legal framework, we support option 2 being a logically centralised model of aggregation. We also believe that optimal implementation can be achieved through a federated data and security approach. This could be done by using a centralised query engine to federate queries to the systems of each of the individual Trade Repositories (TRs), which host a "plugin" that communicates with the central query engine. By adopting this federated approach, the TRs take full ownership and responsibility for normalising and mapping their data to the central harmonised data schema. Their plugin also enforces and implements the security, data privacy and confidentiality rules and then transparently communicates its results to the central query engine.

Our key points are set out below, aligned to the specific questions asked and the structure of the consultation paper.

Legal Considerations – data security, privacy & confidentiality

As noted in the paper, the legal framework to support aggregation of data across different national and regulatory regimes is not yet fully developed. A centralised model of aggregation, as outlined by option 1 would require a complex and adaptable security framework that covers all the nuances of the current legal framework and changes to this framework. We do not believe that this is feasible in a reasonable time frame.

Using option 2: Each TR would take responsibility for embedding their local security, data privacy and confidentiality requirements in the plugin on their local system. This federated approach would ensure that each local jurisdiction, that are best placed to implement these rules given their local knowledge, take full responsibility for encoding them within their local plugin. Any query from the central query engine therefore would obey the rules as encoded through the local plugin and only return results in accordance with those rules. Any data that contravenes local rules can be anonymised or aggregated "on-the-fly" locally to ensure there is full transparency to the central team regarding completeness of the full trade population.

Using option 2: As the legal environment develops, individual plugins would be updated to reflect the changes in international agreements and co-operation arrangements. Crucially, these could be updated independently as agreements are reached on a jurisdiction-by-jurisdiction basis, rather than requiring a global agreement or being held up by other cross-jurisdictional dependencies.

While option 3 could account for the necessary local legal issues, there would be no automated mechanism for making the security filtering that has taken place transparent to the teams responsible for central co-ordination of the various raw data reports from the various TRs.

Legal Considerations - governance framework

We believe that the best approach to setting up a mechanism for running the aggregation mechanism is through a public-private partnership, specifically where a private entity could run the aggregation mechanism and be supervised/overseen by a college of authorities from different jurisdictions. This approach fits best with our recommended federated approach where a private entity is responsible for developing, hosting, running and coordinating activities for the central query platform, under a mandate from and with supervision by the college of authorities. The individual members of the college of authorities would ensure that their plugins on their local infrastructure meets the needs of the overall mandate.

The consultation paper rightly recognises that the aggregation of TR data is a complex task which is why we recommend a federated approach which allows federated entities to manage their local complexities whiles leaving a lighter burden to be managed centrally. In our view, spreading the workload in this manner is likely to result in a more resilient outcome.

Data and Technology Considerations - underlying data

Fundamental to our recommendation of option 2 is our belief that a federated data approach is critical to achieving data integrity. In our view, the closer data is to its original source, the better the quality tends to be. While arguably the original data is the actual trades within each of the underlying reporting entities, the next best alternative is to source directly from the TRs. Using a truly federated approach, the link to the original source data is never lost thus guaranteeing the integrity from that source. Even where anonymisation or aggregation is required to comply with local rules, the plugin on the local TR system keeps full transparency into the underlying trades.

Under option 1, the data is fed into the central store through an Extract-Transform-Load (ETL) process that copies a view of the original data into a new store. This copy-based approach breaks the link with the original data and allows the data providers to relinquish their accountability to this data since they are not responsible for the central store. More importantly, under the current legal framework, the central repository will only be permitted to hold a filtered version of the raw data.

The federated data approach, in Option 2, also allows better data quality, by ensuring that each TR retains responsibility for catering for the different dimensions of data quality, being: Completeness, Accuracy, Timeliness, Consistency, Accessibility and Uniqueness (no Duplication). The central function would still perform data quality checks and other forms of validation, given that they are likely to encounter these as part of their analysis. However these data quality issues should be pushed back to the original TR data sources using a workflow engine that can track and follow up remediation.

The consultation paper correctly identifies standardisation as a being necessary for effective high quality aggregation. Without a shared taxonomy the aggregation of data becomes meaningless. It is necessary to distinguish between standardisation of definition and that of nomenclature since the absence of standard definition would preclude effective aggregation while absence of standard nomenclature would not. Under a federated approach, individual TRs would have responsibility for capturing and populating the data elements required for reporting according to the standardised definitions, however they may use different nomenclature to describe how these data elements are stored locally. Where this occurs, the plugins on the local TR systems would be responsible for mapping the local nomenclature to the global standard nomenclature. The central query engine would provide complete transparency by showing both the globalised view as well as the local mappings. This would allow meaningful communications in the event that the central function requires clarification of data from a local TR. For example, if a data element such as Market Value is defined differently by separate TRs, one with accrued interest and one without, then meaningful aggregation is compromised, however if one TR calls it Market Value while another call it Present Value, aggregation is still meaningful, but it would be useful to retain transparency on the different nomenclature.

It is essential that any aggregation mechanism is able to adapt to changing standards as well as different rates of adoptions of the TRs to these standards. This implies that the technology supporting the aggregation mechanism has the required flexibility to adapt to this level of change. We believe that option 2 provides the best mechanism of providing the necessary flexibility as discussed in the technological arrangements below.

A centralised approach, as suggested by option 1, is more suited to an environment where standards are known and don't change significantly over time. This allows the aggregation mechanism can be designed and built accordingly. Our experience is that large centralised warehouses often fail not due to technical limitation on the size or volume of data but rather on the inability to adapt the data schema due to changes in the environment or requirements.

Data and Technology Considerations – technological arrangements

Before considering each of the criteria of an effective technological arrangement it is worth highlighting a point raised in the AFSG outreach workshop where some participants stated that they would not recommend pre-aggregation of the data and advised that the aggregation mechanism should use the raw transactional data. We agree with this view as it provides the level of transparency required. Given this view, the technological arrangements need to be assessed against this context.

Scalability – we believe option 2 provides the best alternative for scalability given the volume of data involved (many billions of rows of raw transactional data). Whilst option 1 may provide the technical capabilities of storing this amount of data, we believe the timeliness of transferring large volumes of data through an ETL process would compromise the responsiveness of a centralised platform. Under option 2, there is no need for regular data transfers and data is required only on an as-needed basis. Furthermore, using a federated data approach, the central query engine can federate queries out to underlying TR data stores where the data requirements can be filtered accordingly to ensure that only the minimum data results are required to be physically moved. Although we still anticipate that this will require technology to be capable of handling many billions of rows of data in each TR, this could be provided using high performance caching within TR plugins. Option 3 requires significant human intervention and therefore we believe it does not meet the scalability requirements.

Flexibility – as noted above, the amount of expected change over time in both the TRs and global standards means that the technology must be agile enough to cope with this change. A data warehouse approach under option 1 would not be able to respond to schema changes quickly enough and the testing and co-ordination of changes to data feeds would compromise the flexibility of the aggregation mechanism. We believe that option 2 provides the best approach to handling change. It is possible for a central query engine to provide a virtual data warehouse (the data continues to reside in the federated data stores). This virtual data warehouse would dynamically adapt to change without the need for long and expensive schema re-designs necessary in a physical data warehouse. This would allow the TRs to change independently of each other with less global coordination.

Using a virtual data warehouse also allows flexibility in the initial build making it possible to implement functionality incrementally without an apriori knowledge of the final schema and requirements. For example a limited set of data elements could be provided across all TRs and other elements added incrementally.

This approach is ideally suited to the FSB requirements where the nature of the queries is not fully known in advance and is likely to undergo rapid evolution as the external market environment changes and as users become more sophisticated in their understanding of the system capabilities and data available.

Business continuity – Under all options the TRs would continue to be responsible for their own business continuity. For option 1 the physical data warehouse would need to be duplicated using standard techniques. Option 2 would use a similar approach perhaps with more distributed capacity. This would be both more cost effective, given the much lighter infrastructure required, and more resilient.

Data security – the advantages of a federated approach for data security have been covered above, however from a technology perspective we believe there needs to be a two-tier security model to cater for the FSB requirements. First, the plugins at each TR should by default inherit the security model of their underlying infrastructure which caters for the local security requirements and second, an additional layer of security on the central query platform should provide restricted access for individuals based upon their level of authority to view data.

Criteria to assess aggregation options

We agree with the criteria identified in Chapter 6 used to assess the aggregation options. However we would argue that these criteria should be evaluated on the both tangible evidence and also the less tangible elements such effective use of the data and accountability.

In evaluating the uses of data, it is necessary to consider whether data captured is used effectively or whether it is better to have a mechanism that can provide access based upon specific needs and requests. If data is visibly being used to good effect, we believe that organisations are more likely to give it the focus it demands.

In evaluating data integrity it is necessary to consider the impact of accountability on each option. In our view, a mechanism where individual TRs provide data feeds to a central warehouse can create an environment where they no longer feel accountable for that data. In a federated model the approach taken is to create a central query that is federated out to a series of sub-queries into TR data stores. Using this approach the TRs retain accountability since it is their data that is being queried and is not copied so as to prevent the likelihood of it being altered or adjusted further downstream.

Finally, we note that cost does not feature as a criterion for evaluation. We would recommend that this be included as a factor.

Other aggregation options

We do not believe there are any further aggregation options that could be considered. However under option 2 we do make a distinction between a fully federated approach where data continues to reside locally and only final results of federated queries are passed back centrally versus a pseudo-federated approach where data from TRs is pulled into a centralised in-memory cache from where analysis and queries can be performed. Our recommendations are based upon a fully federated approach.
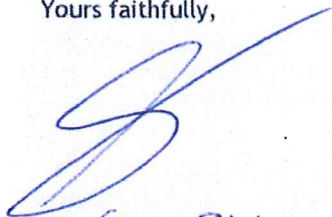
Other uses for aggregated data

Having a mechanism available to query and aggregate trade data is extremely valuable from a regulatory perspective beyond the original mandate of financial stability. Individual regulators and authorities have many different mandates which could be fulfilled by such functionality such as financial conduct and prudential mandates beyond financial stability.

In addition, financial services companies who are providing the underlying data to the TRs could also benefit if the central aggregation mechanism were able to leverage the use of that data to provide services back to the contributors. For example, as noted in the AFSG outreach workshop, meaningful analytical information is derived from the ability to understand how changes in risk factors affect value, income, liquidity and stress tests. Such analysis starts with the underlying trade information.

It is possible, with participation from a private entity running the central utility, to provide additional functionality which can leverage the trade information. An example could be to provide the stress scenarios and valuation results in a consistent way. This could also be coordinated with regulators as part of the many stress testing activities already required. This additional functionality could be provided with a market data provider who already has access to the data elements required for estimating the changes in underlying risk factors.

We trust you find our comments helpful and would be happy to discuss further any part of our response, or to participate in further outreach workshops or committees that may be held on this topic.

Yours faithfully,

Gavin Slater
Director